

Zadanie 1. Analiza pojedynczego zdania ----

Wczytaj dane tekstowe

```
text <- "And so even though we face the difficulties of today and tomorrow, I still have a dream."
```

```
text
```

Sprawdź częstotści słów za pomocą pakietu qdap

```
install.packages("qdap")
```

```
library(qdap)
```

```
freq_terms(text)
```

Zapisz najczęściej występujące terminy w ramce danych

```
frequent_terms <- freq_terms(text)
```

```
frequent_terms
```

Wizualizacja najczęściej występujących terminów

```
plot(frequent_terms)
```

UWAGA

Słowa nie są wymienione w takiej kolejności, w jakiej występują w zdaniu

są prezentowane w porządku alfabetycznym.

Takie podejście nazywa się Bag of Words (torba słów).

Inne możliwości pakietu qdap

```
?freq_terms
```

```
# Wizualizacja za pomocą ggplot2
```

```
library(ggplot2)
```

```
ggplot(frequent_terms, aes(x = WORD, y = FREQ)) +  
  geom_bar(stat = "identity", fill = "skyblue") +  
  labs(x = "Słowo", y = "Częstość") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +  
  ggtitle("Wykres częstości słów")
```

```
ggplot(frequent_terms, aes(y = WORD, x = FREQ)) +  
  geom_bar(stat = "identity", fill = "skyblue") +  
  labs(x = "Słowo", y = "Częstość") +  
  theme(axis.text.x = element_text(angle = 0, hjust = 1)) +  
  ggtitle("Wykres częstości słów")
```

```
# Bardziej atrakcyjna wizualizacja
```

```
ggplot(frequent_terms, aes(x = FREQ, y = reorder(WORD, FREQ))) +  
  geom_bar(stat = "identity", fill = "skyblue", color = "darkblue", alpha = 0.8) +  
  labs(x = "Częstość", y = "Słowo") +  
  ggtitle("Wykres częstości słów") +  
  theme_minimal() +  
  theme(axis.text.y = element_text(size = 10), # Dostosowanie rozmiaru czcionki etykiet na osi Y  
        plot.title = element_text(hjust = 0.5, size = 16, face = "bold"), # Wyśrodkowanie i stylizacja tytułu wykresu  
        panel.grid.major.y = element_blank(), # Usunięcie głównych linii siatki poziomej  
        panel.grid.minor.y = element_blank(), # Usunięcie mniejszych linii siatki poziomej  
        axis.line = element_line(color = "black")) # Dostosowanie linii osi
```

```
# Stopwords (stop słowa – słowa do usunięcia)
```

```
# Najczęściej występujące 25, 100 i 200 słów
```

```
Top25Words
```

```
Top100Words
```

```
Top200Words
```

```
# Usunięcie stop słów
```

```
frequent_terms2 <- freq_terms(text, stopwords = Top25Words)
```

```
frequent_terms3 <- freq_terms(text, stopwords = Top100Words)
```

```
frequent_terms4 <- freq_terms(text, stopwords = Top200Words)
```

```
plot(frequent_terms4)
```

```
# Zadanie 2. Analiza całego akapitu ----
```

```
# Wczytaj dane tekstowe
```

```
text <- "And so even though we face the difficulties of today and tomorrow, I still have a dream. It is a dream deeply rooted in the American dream."
```

```
text
```

```
frequent_terms <- freq_terms(text)
```

```
frequent_terms
```

```
frequent_terms <- freq_terms(text, stopwords = Top200Words)
```

```
plot(frequent_terms)
```